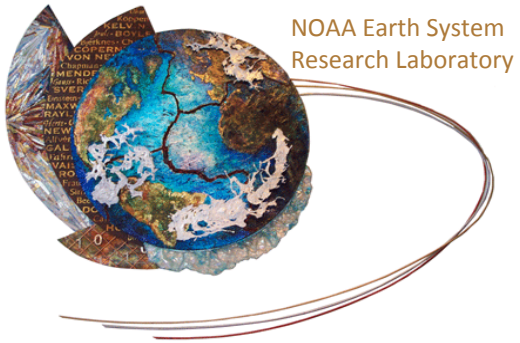
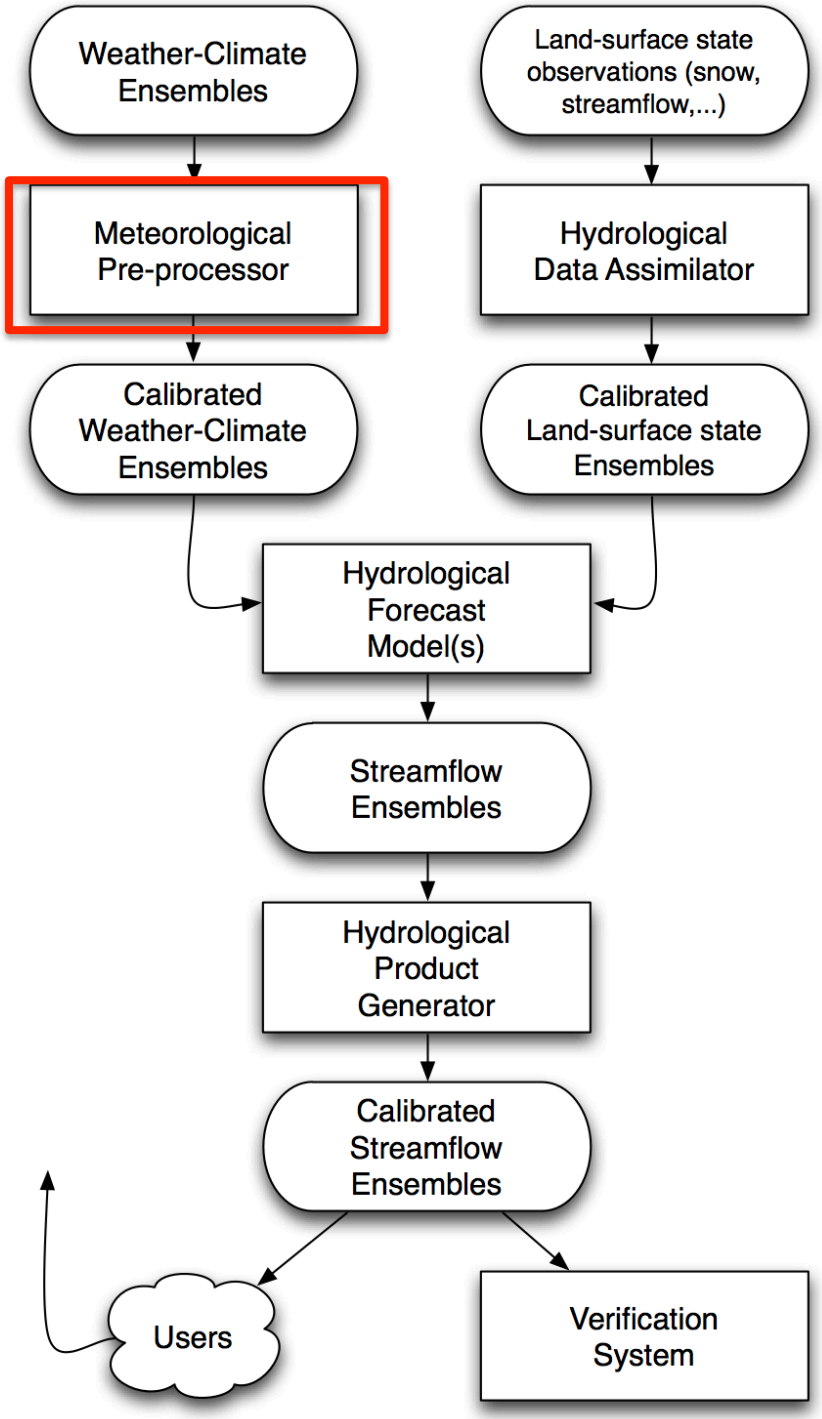




# A new multi-decadal global ensemble reforecast data set using the NCEP GFS

Tom Hamill, Jeff Whitaker,

and Gary Bates

NOAA ESRL Physical Sciences Division




# HEPEX's envisioned hydrologic ensemble prediction system

Our focus is the box in red.
Why meteorological pre-processing?
What data sets are needed?

from Schaake et al., *BAMS*, 2007.
see also DOI: 10.1002/asl.267

# Why meteorological pre-processing?

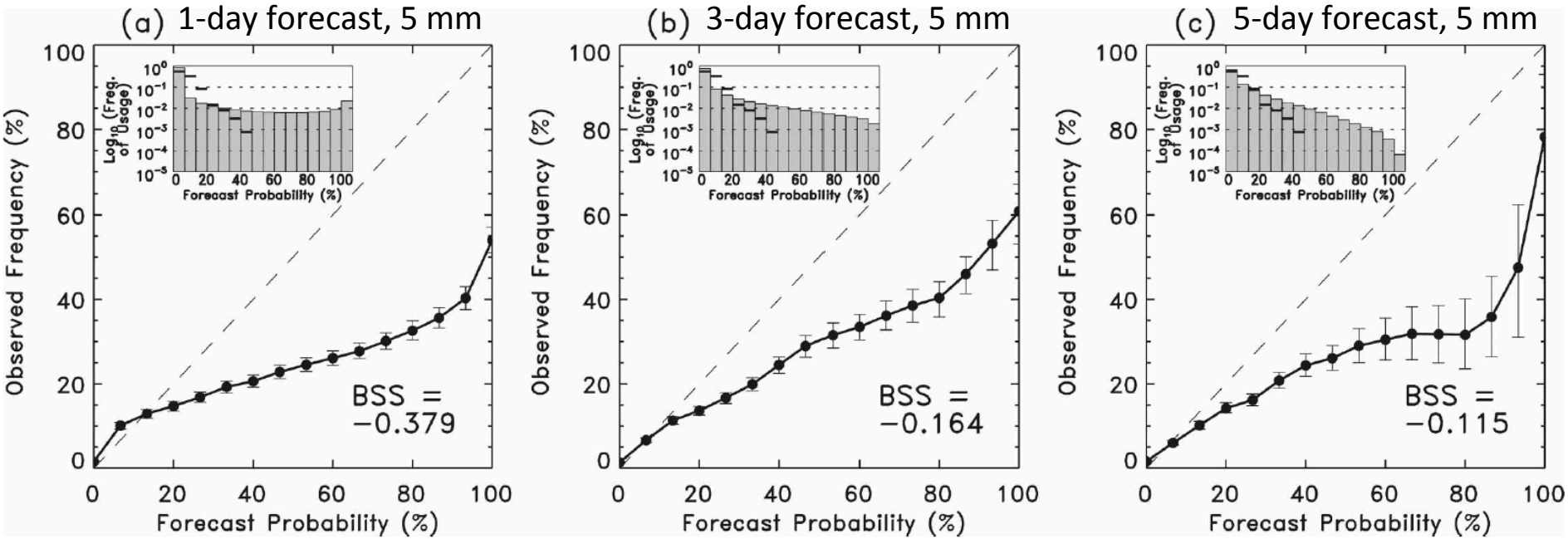
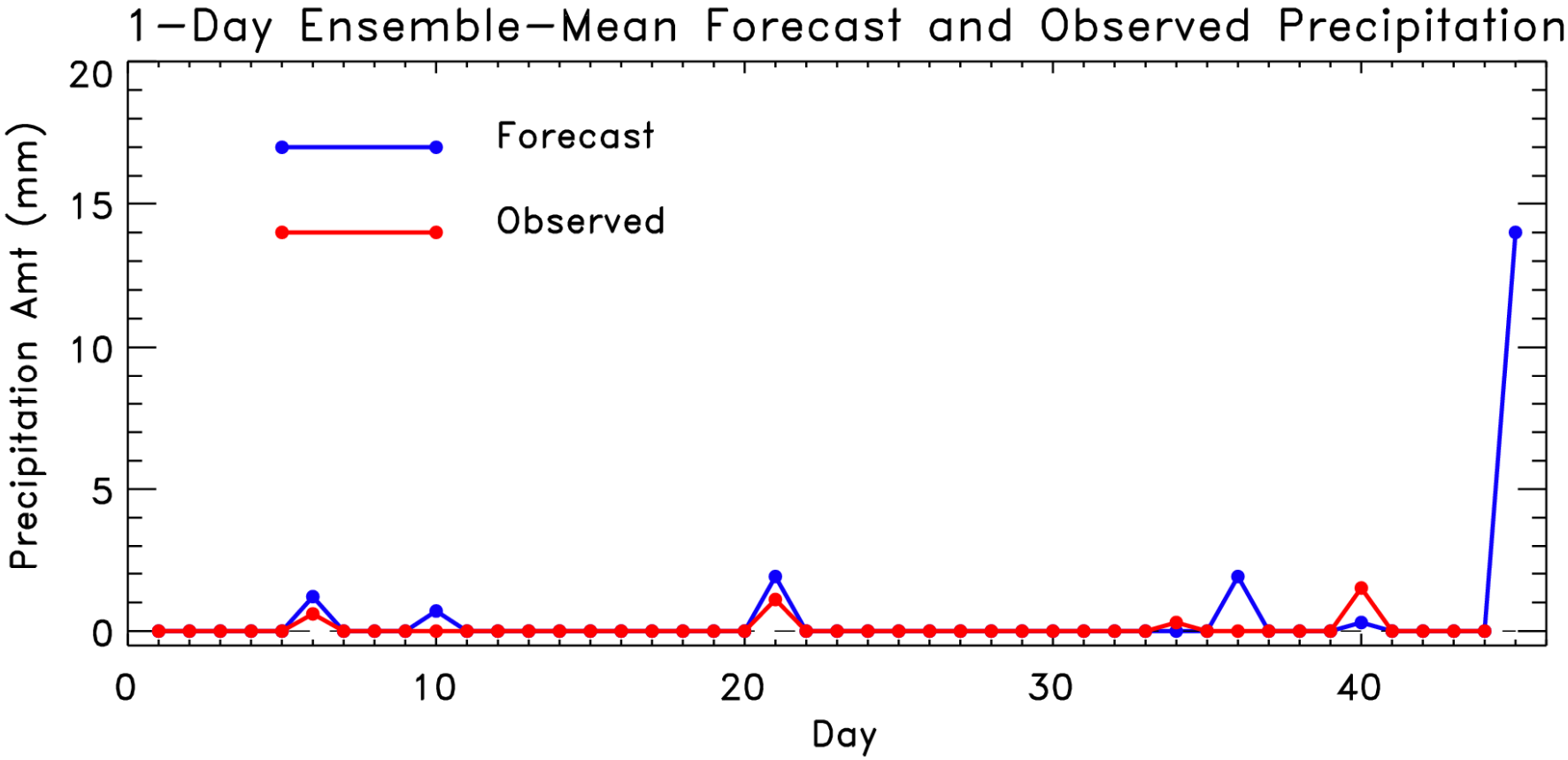


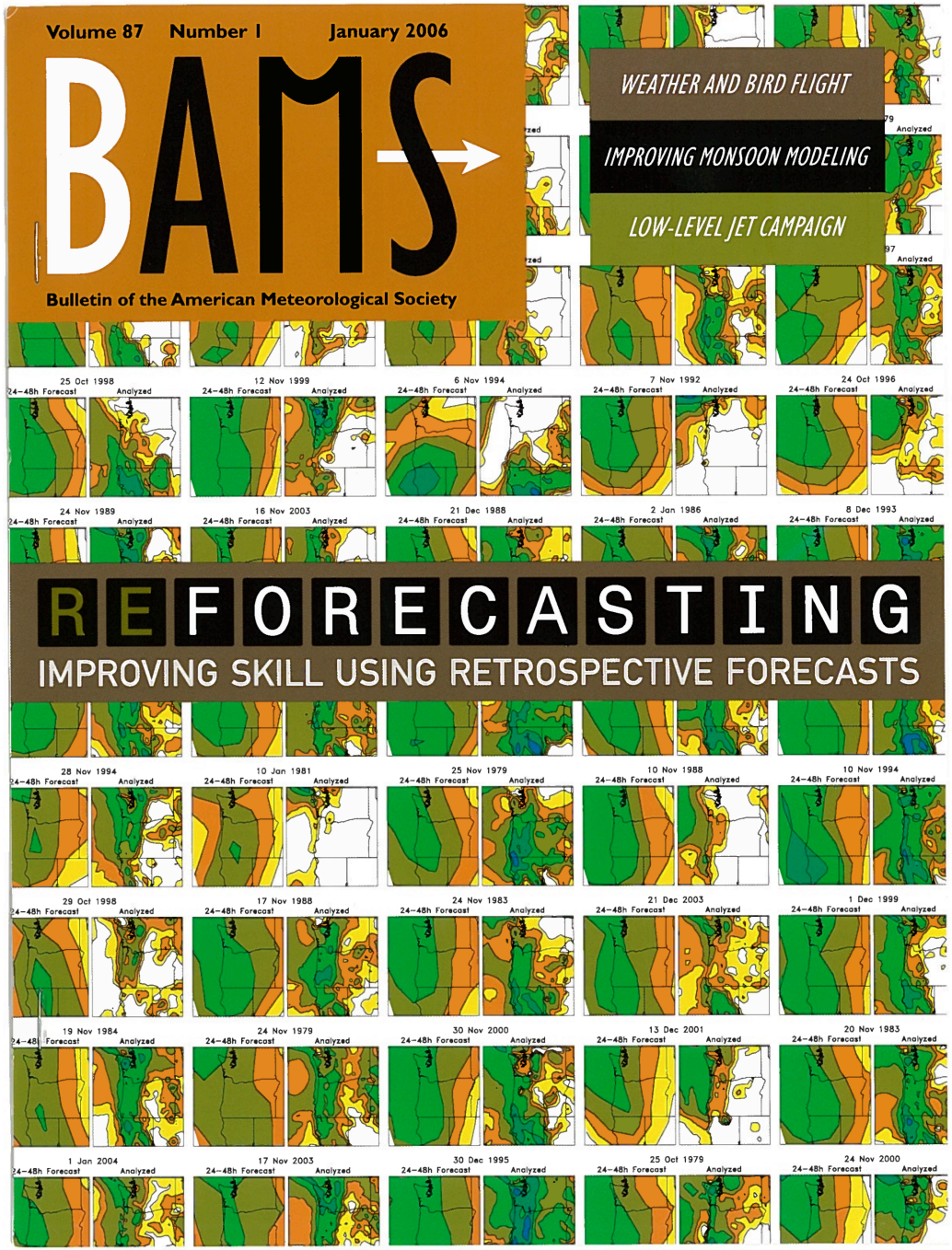

FIG. 4. Reliability of 5-mm ECMWF raw forecasts at (a) 1-, (b) 3-, and (c) 5-day leads. Plotted confidence intervals show the 5th and 95th percentiles as determined through block bootstrap resampling techniques. The inset histogram denotes frequency of forecast usage of each probability bin. Solid lines plotted on the histogram denote the climatological frequency of usage of each probability bin.

Even ensemble prediction systems such as the ECMWF system typically do not yet provide highly reliable probabilities. Model imperfections, limited size ensemble, imperfect method of initialization (and imperfect observations) all contribute.

# Making reliable forecasts for rare events complicated w/o large training sample




A heavy precip event like the one today are the ones you care about the most. How to calibrate today's forecast given past short sample of forecasts and observations?



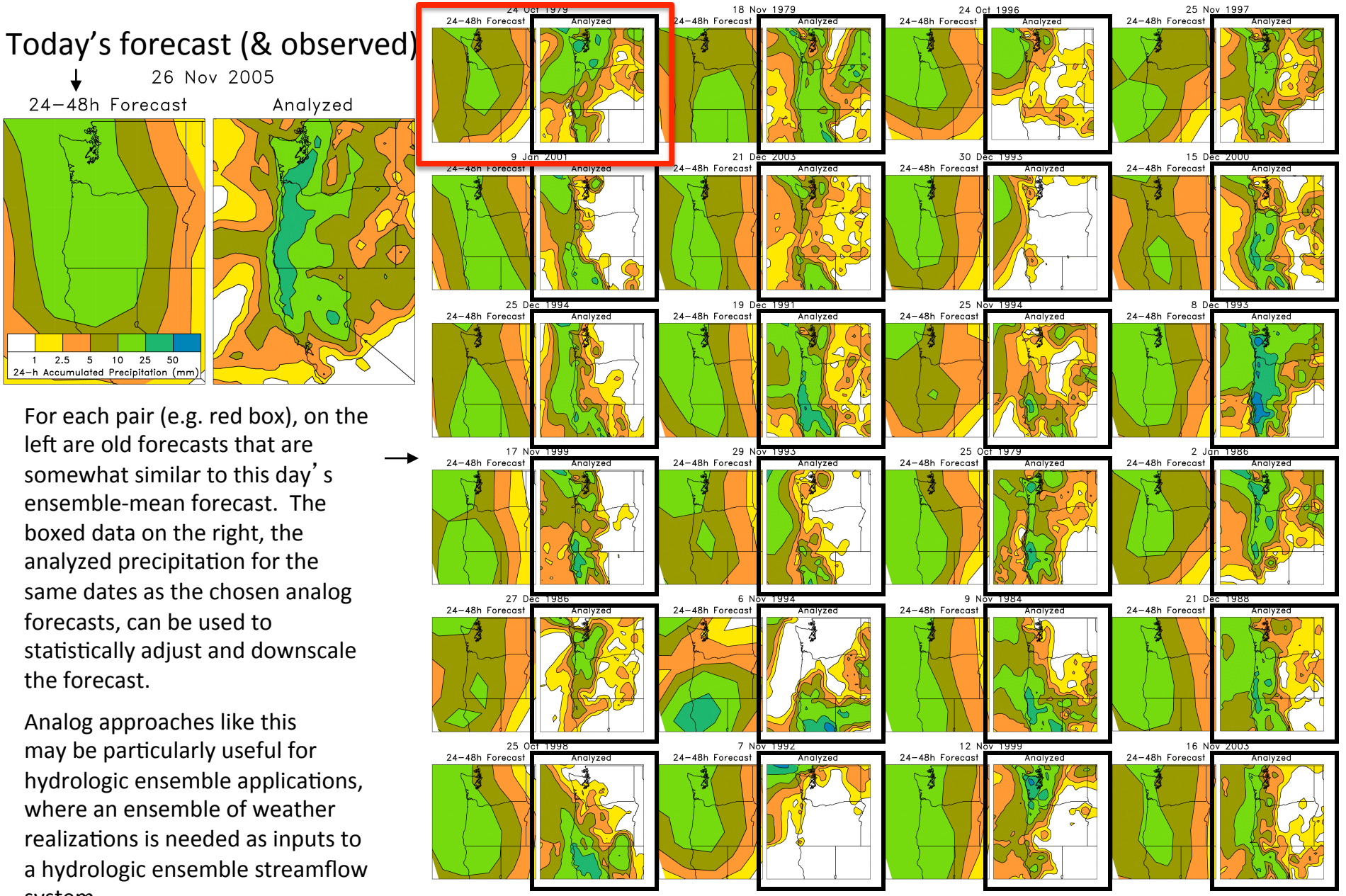

We created a "reforecast" data set ~ 8 years ago with a 1998 version of the NCEP GFS ensemble prediction system.

15 members, 15-day forecasts, every day 1979 to current.

However:

(1) only a few select fields archived
(2) models improved greatly since then.

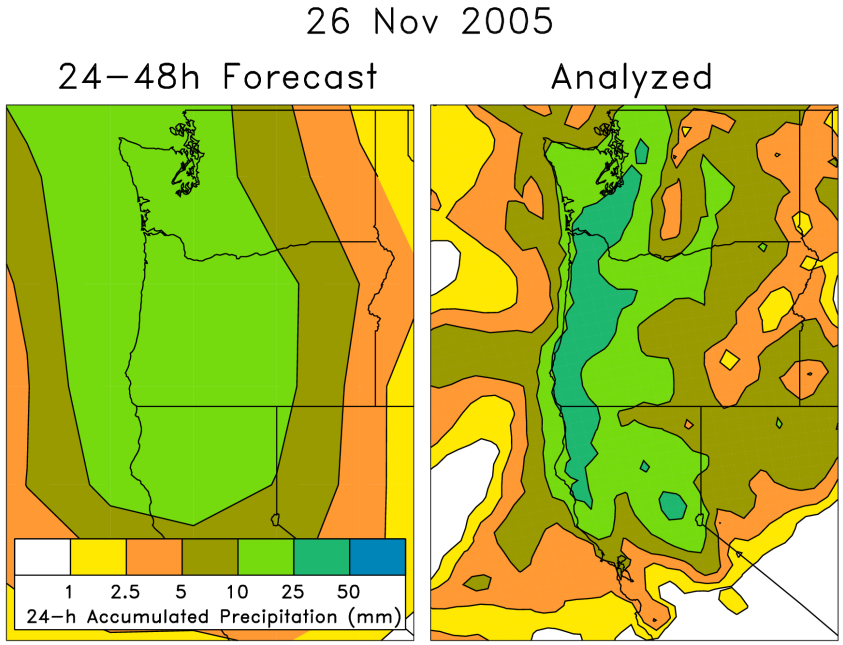
# An example of a statistical correction technique using those reforecasts



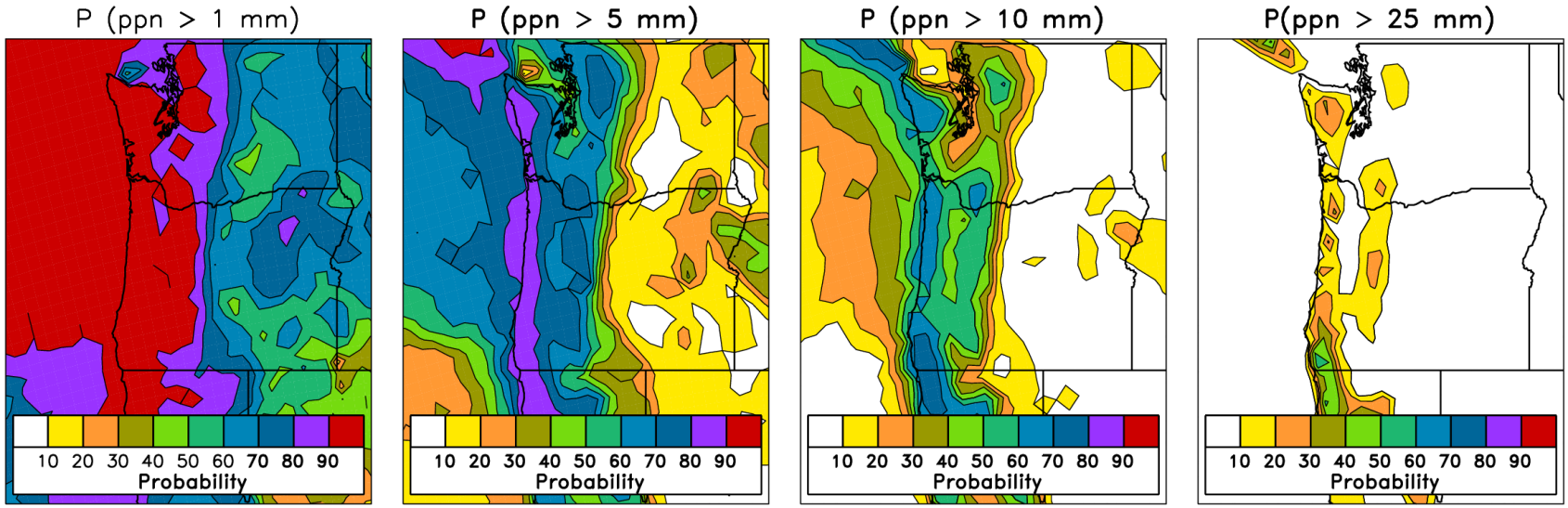

For each pair (e.g. red box), on the left are old forecasts that are somewhat similar to this day's ensemble-mean forecast. The boxed data on the right, the analyzed precipitation for the same dates as the chosen analog forecasts, can be used to statistically adjust and downscale the forecast.

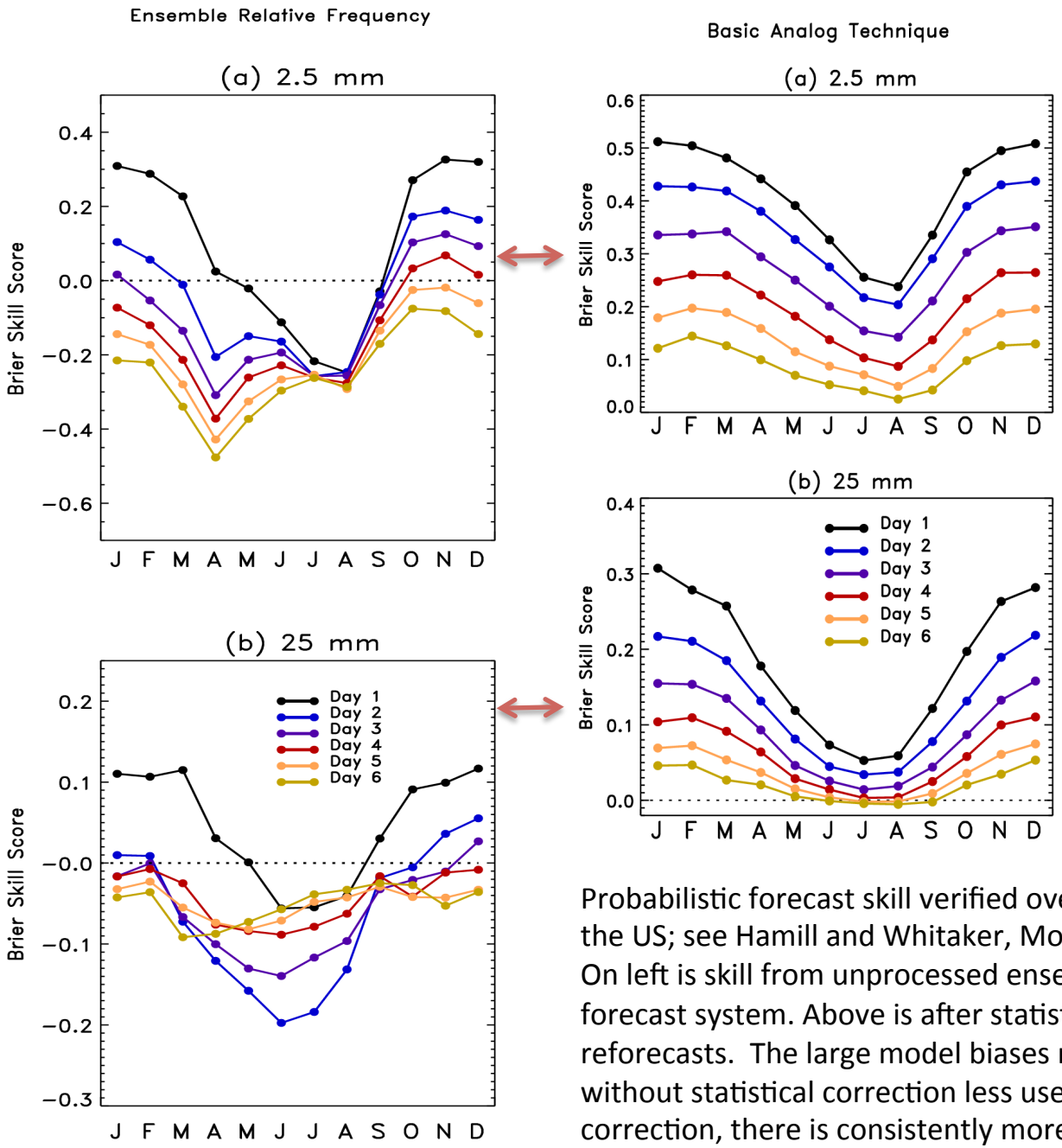
Analog approaches like this may be particularly useful for hydrologic ensemble applications, where an ensemble of weather realizations is needed as inputs to a hydrologic ensemble streamflow system.

# Downscaled analog probability forecasts




Statistically adjusted analog forecast probabilities



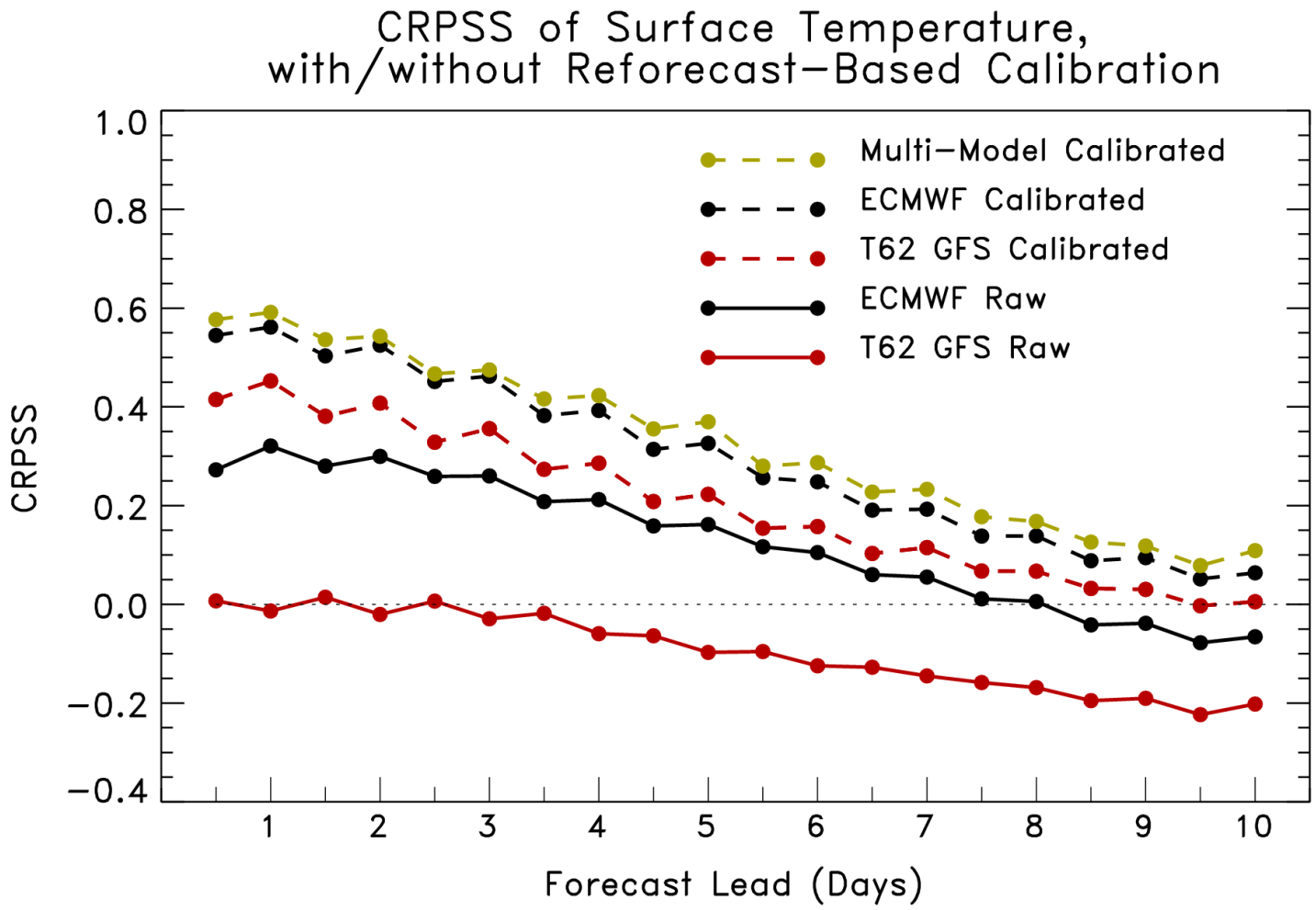



# Does this improve the forecast?

Yes, statistical correction techniques like these have produced dramatically improved forecast guidance.

Probabilistic forecast skill verified over 25 years of forecasts across the US; see Hamill and Whitaker, Monthly Weather Review, 2006. On left is skill from unprocessed ensemble from 1998 GFS ensemble forecast system. Above is after statistical post-processing using reforecasts. The large model biases make the ensemble forecasts without statistical correction less useful than climatology. After correction, there is consistently more skill than climatology.

# Reforecast skill improved with newer models




Our 1998 T62 GFS reforecast and ECMWF's 2005 T255 reforecast

# Proposal

(supercomputer grants to renewable-energy proposals)


- Give us the supercomputer cycles, and we'll generate a new reforecast database with a version of the GFS that will be operational in 2011.
- We'll demonstrate long-lead, probabilistic solar-energy potential, wind-energy potential, hydro forecasts.
- The database will be freely available for your use.

# What we have been granted by DOE, and what's expected of us

- 14.5 M CPU hours (CPUh) on "Franklin" supercomputer at Lawrence Berkeley lab, all to be used before June 2011.
  - some of those cycles (10%) needed for generating initial conditions, too.
- We are expected to make this data set available to the community by late 2011, and to demo those experimental wind, solar, and hydro forecasts.

# Principles

- Reforecasts will be computed with a (smaller-ensemble) version of the GEFS that will be operational in 2011.
- We hope that GEFS will remain in this configuration, or will be changed only slightly, for several years thereafter.
- Once GEFS changes, either EMC or ESRL will continue to run the reforecast version until a next-generation reforecast is in place.

# Anticipated configuration

- At 00Z, full 10-member forecast, every day, for last 30 years out to 16 days. T254L42 to day 8, T190L42 from days 7.5 to day 16. CFSR initial conditions (GSI). Mimics operational configuration in late 2011.
- If, at the end of DOE grant, there are extra CPU cycles, they may be devoted to:
  - running some 45-day reforecasts at T126 for the sake of comparison against CFSR reforecasts, with coupled ocean model (it'd be nice to have a few spare cycles to do this).
  - adding a small number of reforecasts for the 12Z cycle.

# Storage of data

- Storing of "important" agreed-upon subset of data ~= 130 TB. Which fields described in a subsequent slide.
  - ESRL / PSD is purchasing ~ 200 TB of storage and server capability for this data set. Cost ~ $160K for hardware. Hopefully procurement finished in several months.
  - Will design software to serve this out to you in several manners (http, ftp, OPeNDAP, etc.).
  - Back this up to tape.
- Storing full 00Z reforecasts and initial conditions ~=778 TB.
  - Useful for LBC's to run regional reforecasts, or if more fields added to "important" subset.
  - NCDC / NGDC our only realistic storage solution. They may want funds we don't have.
  - Likely we will likely save the initial conditions but not the forecasts.

# Proposed fields for “fast” archive

- Mean and every member
- “Fast” archive will be on disk, readily accessible
- Mandatory level data:
  - Geopotential height, temperature, u, v, at 1000, 925, 850, 700, 500, 300, 250, 200, 100, 50, and 10 hPa.
  - Specific humidity at 1000, 925, 850, 700, 500, 300, 250, 200
- PV (K $m^2$ $kg^{-1}$ $s^{-1}$ ) on θ = 320K surface.
- Wind components, potential temperature on 2 PVU surface.

# Fixed fields to save once

- field capacity
- wilting point
- land-sea mask
- terrain height

# Proposed single-level fields for “fast” archive

- Surface pressure (Pa)
- Sea-level pressure (Pa)
- Surface (2-m) temperature (K)
- Skin temperature (K)
- Maximum temperature since last storage time (K)
- Minimum temperature since last storage time (K)
- Soil temperature (0-10 cm; K)
- Volumetric soil moisture content (proportion, 0-10 cm) –
- Total accumulated precipitation since beginning of integration (kg/m$^2$)
- Precipitable water (kg/m$^2$, vapor only, no condensate)
- Specific humidity at 2-m AGL (kg/kg; instantaneous) –
- Water equivalent of accumulated snow depth (kg/m$^2$) –
- CAPE (J/kg)
- CIN (J/kg)
- Total cloud cover (%)
- 10-m u- and v-wind component (m/s)
- 80-m u- and v-wind component (m/s)
- Sunshine duration (min)
- Snow depth water equivalent (kg/m$^2$)
- Runoff
- Solid precipitation
- Liquid precipitation
- Vertical velocity (850 hPa)
- Geopotential height of surface
- Wind power (=windspeed$^3$ at 80 m*density)

# Proposed fields for “fast” archive

- Fluxes ($W/m^2$ ;  average since last archive time)
  - sensible heat net flux at surface
  - latent heat net flux at surface
  - downward long-wave radiation flux at surface
  - upward long-wave radiation flux at surface
  - upward short-wave radiation at surface
  - downward short-wave radiation flux at surface
  - upward long-wave radiation at nominal top
  - ground heat flux.

# Where we are now

- 200 TB storage & server procurement underway, delivered in months.
- Control run mostly done, to 8 days. Awaiting remaining 25% or so of CFSR initial conditions from NCDC.
- Software for generating perturbations mostly ported.
- Having trouble getting "throughput" on DOE computers.
  - Will get some boosting in priority, but…
  - Still may not be able to use all cycles by June, so hopefully carry them over and complete reforecast in June-December timeframe.

# Prepare!

- Should you wish to get some of this data, be aware that storing even a subset of this data will probably require several TB, e.g., storing all fields in a rectangle covering N America will require 15 TB storage.

# Conclusions

- Complete reforecast database with modern model has long been on hydrologists' wish list.
- Within a year, we expect to have that for you.
- We intend to make it easy for you to access this data for your R&D.
- We hope this data set will spur many advances in hydrologic prediction and ensemble weather forecasting.